DOCUMENT RESUME

ED 310 141 TM 013 724

AUTHOR Plomp, Tjeerd; van der Meer, Adri
TITLE A Method for a Retrospective Analysis of Course Objectives: Have Pursued Objectives in Fact Been Attained? Twente Educational Report Number 7.
INSTITUTION Twente Univ., Enschede (Netherlands). Dept. of Education.
PUB DATE Feb 79
NOTE 41p.; Paper presented at the Conference of the European Association for Research and Development of Higher Education (3rd, Klagenfurt, Austria, January 1979).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS College Curriculum; *College Freshmen; College Mathematics; *Course Evaluation; *Educational Objectives; Engineering Education; *Evaluation Methods; Foreign Countries; Higher Education; *Instructional Effectiveness; Outcomes of Education
IDENTIFIERS *Retrospective Analysis

ABSTRACT

A method pertaining to the identification and analysis of course objectives is discussed. A framework is developed by which post facto objectives can be determined and students' attainment of the objectives can be assessed. The method can also be used for examining the quality of instruction. Using this method, it is possible to determine mathematics course objectives and the attainment of these objectives by the group of students who passed the course. Subjects of the study were 816 first-year engineering students, enrolled in a university program during the 1970-1973 period. The students were administered quarterly open-question examinations during the first three quarters of each year. The retrospective analysis data were collected in 1976. The exploratory character of this research suggests that conclusions from the data must be considered merely as illustrations of the type of questions that can be analyzed within the proposed framework. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED310141

TWENTE EDUCATIONAL REPORT NUMBER 7

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. NELISSEN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A METHOD FOR A RETROSPECTIVE ANALYSIS OF COURSE OBJECTIVES: HAVE PURSUED OBJECTIVES IN FACT BEEN ATTAINED?

Tj. Plomp

A. van der Meer

TWENTE UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF APPLIED EDUCATION
P.O. BOX 217
7500 AE ENSCHEDE THE NETHERLANDS

TM013724

BEST COPY AVAILABLE

Uitgave : Onderafdeling Toegepaste Onderwijskunde, THT
Typewerk : Josë Oberjë
Reproduktie : Centrale Reproduktieafdeling, THT
Oplage : 100

A method for a retrospective analysis of course objectives: Have pursued objectives in fact been attained?

by

Tjeerd Plomp (Department of Educational Technology)
Adri van der Meer (Department of Applied Mathematics)
Twente University of Technology,
P.O. Box 217, 7500 AE Enschede, The Netherlands

Presented at the third conference of the European Association for Research and Development of Higher Education, Klagenfurt (Austria), January 1979.

Twente Educational Report No. 7. Enschede, The Netherlands: Twente University of Technology, Department of Applied Education, February 1979.

1. Summary

In higher education, course objectives are seldom explicitly stated. Consequently, it is almost impossible to indicate which objectives have been attained by the students at the end of a course. Thus, there is often a need to establish, post facto, which objectives were pursued by the teachers and attained by the students. This need is especially important where a new instructional method is weighed against the usual one or where the identification of critical impact points for improving an existing instructional method is the intent.

An analysis of course examinations over several years can provide good insight about the objectives teachers have pursued, because the examination items can be considered operationalizations of the pursued objectives. And, by analyzing achievement on the examination problems, it is possible to determine which objectives the students have attained.

In § 3 of this paper a method pretaining to objective indentification and analysis will be discussed. Thus a framework will be developed by which post facto objectives can be determined and students' attainment of the objectives can be assessed. The method can also be used for examining the quality of instruction (see § 5).

Using this method, we are able to determine mathematics course objectives and attainment of the objectives by the group of students who passed the course (see § 4). The results of this study are based on first-year engineering students in the period 1970-1973, using "open question" examination formats. The explorative

character of this research suggests that conclusions from the data have to be considered illustrations of the type of questions that can be analysed within the proposed framework.

2. Research-questions.

In this paper, we discuss only a few of the research questions that could be discussed in the context of a retrospective analysis of implicit objectives.

The central question is: what objectives were attained by the students, who passed the course? Several questions can be distinguished form this general question:

1. How can the pursued objectives be made explicit, so that they can be discussed as well on content levels as on behavioral levels?
2. What were the objectives (on both levels) pursued in the mathematics courses?
3. What were the objectives (on both levels) the passed students attained?
4. Were in the examinations the number of items, which are operationalizations of the objectives attained by the group of passed students, such a proportion of the respective examination, that based on the scores on these items a passmark could obtained? If not, is it possible to characterize the not-attained objectives, of which then must be concluded that the passed students have not or only partially attained them?

The relevance of these research questions comes from the situation that passed students (i.e. students with a pass mark) are considered to have attained course objectives, while often in practice evidence about this assumption is not considered.

Ideally, questions pertaining to the attainment of objectives should be answered within a research design in which all the items from different examinations are connected by calibrating them on the same scale. For this it is necessary either to present some items to persons from different groups or to calibrate afterwards all the examination items on one group. In practice this requirement can not be met, so a retrospective analysis, in which plausible assumptions are made, is needed to answer the above questions.

3. Method of research

In this section we introduce the method of research which will be applied in the next section as a procedure by which the research questions can be answered.

Considering the first question, we propose a two-dimensional grid fitting the situation of examinations based on implicit objectives (§ 3.1.). After having classified the examination items within the grid, question 2 can be answered (§ 4.4.). To address questions 3 and 4, it is necessary to introduce in § 3.2. such constructs as: (1) item in an "open question" examination, (2) the mastery of an item by a student, and (3) measure for attainment of an objective.

The method for conducting the retrospective analysis of course objectives is based upon the analysis of examinations results, which may be influenced by several sources of variation. These sources and some assumptions are discussed in § 3.3., the method of analysis in § 3.4.

3.1. Grid of objectives

The items or problems in course examinations can be considered operationalizations of the pursued objectives. Each item is the operationalization of a particular objective, that may be described on two dimensions: (a) content and (b) behavior. Content refers to specific substantive elements in the body of knowledge. If the content levels are formulated too specifically, items may be a composite of several objectives. Besides, then a too large number of cells should be generated so that per cell not a meaningful number of items will be obtained. Therefore, the content levels must be formulated broadly enough to enable an unambigious classification and a meaningful interpretation. This can be done by using general terms, such as chapter titles from text books or main topics from the subject field. For the behavioral dimension, several classifications are available in the literature (e.g. the taxonomy of Bloom (1956) for the cognitive domain or its elaboration for mathematics by Wilson (1971). However, it is essential that the intellectual skill domain is as exact a classification of the measured

behaviors as possible. Since the analysis of objectives starts with examination problems, then the nature of these questions must be reflected in the behavioral levels. Therefore, we have chosen as a classification system the main categories of Polya (1957), who distinghuishes between "problems to find" and "problems to prove". In each of the two categories some levels are defined, reflecting the different levels of difficulty of the examination problems (see § 4.2.)
The cells of the grid and the row and columm totals should be studied as starting points of the analysis. Since examinations are samples of content and intellectual skills, it is much more likely that the cells of the grid will be representative of the course objectives if a number of examinations is aggregated. Then, not only a good impression of what objectives were pursued can be obtained from the grid, but also from the number of items per cell and per marginal total an impression of the relative importance of the respective objectives. A caution should be noted. Empty or sparingly filled cells do not necessary mean that the objectives for the cells were not pursued; such cells may appear e.g. if there is some hierarchy within the colums or the rows, so that some objectives are implicitly included in other objectives.

3.2. Definitions

An item or problem is the smallest unit in an examination to which a separate score will be given by the corrector and which

can be solved independently from the other problems in the examination. That means that so called 'pile'-questions (i.e. questions of which the solution is dependent on the results of one or more preceeding questions) will be considered as one item or problem. The m items in a cell are noted as $v_i (i = 1, 2, \ldots, m)$.

Student j has mastered item v_i if his score on v_i is more than a certain fraction of the maximum score S_i; in other words if his score S_{ij} on v_i applies to: $S_{ij} \geq \alpha S_i$ with $0 \leq \alpha \leq 1$.
In every application of the method, the value of the parameter α must be chosen. This choice will depend on 'local' factors like the character of the objectives of a particular course or judgements about the appropriate level of mastery.

Attainment of an objective will be defined with the help of the definition of mastery of an item by single students.
By introducing the parameter α, the score of a student on an item is dichotomized. This can be made explicit by stating a 1 if $S_{ij}/S_i \geq \alpha$ and a o if the reverse holds.
A measure of attainment of an objective operationalized in the item v_i is the ratio:

$$p_i = \frac{\text{number of students with a 1 on } v_i}{\text{total number of students on } v_i}$$

3.3. Assumptions and sources of variation.

Students differ in their capability for mathematics. If we indicate with θ the continuum underlying the ability of students to solve mathematics problems, then every student has a position on this continuum, and it is meaningful to assume a probability distribution $g(\theta)$ for the group students.

The quantity p_i, introduced above, is an estimation based upon one item of the stochastic variable π, the true measure of attainment of the objective represented by the cell of the grid from which the item is drawn.

Consider a cell of the grid with items v_i with p_i ($i = 1, 2, \ldots, m$). Generally, the p_i ($i = 1, 2, \ldots, m$) in a cell will differ. Apart from influences by chance, this should not be the case because all the items in the cell are operationalizations of one objective. Important questions are: (a) how do we explain this variation and (b) how do we combine the information from all the items in a cell and to interpret this on one scale as a measure for the attainment of the objective represented by that cell. Three sources of variation can be considered: (a) students, (b) instruction and (c) items. Students differ in ability, preparation, attitude, etc. The quality of instruction is influenced by different teachers personalities, teaching-styles, instructional procedures, etc. Since these two types of variation can be considerable and will influence the shape of $g(\theta)$ it is important to aggregate different course examinations where as little variation has occurred as

possible. We assume that the combined influence of both sources is constant over the considered period of time and that $g(\theta)$ has the standard normal distribution $\phi(\theta)$ (Assumption 1). The first part of this assumption is plausible in situations in which, over the period of time during which a course will be studied, nothing has been changed with respect to (a) the entry requirements, (b) the preceeding instruction, (c) the group of executive teachers, (d) the instructional procedures, (e) the basic textbook, etc.

In the third source of variation, items may differ in difficulty and use of enabling skills, etc. Moreover, in the situation of the retrospective analysis, the items come from different examinations, taken by different groups of students (from the same population). Considering the practice of aggregating examinations, it follows from the first assumption that differences in p-values in a cell are due to differences between items. Thus, for a cell of the grid the $p_i (i = 1, 2, \ldots, m)$ not only pertain to a measure of attainment of the objective as represented in the respective items v_i $(i = 1, 2, \ldots, m)$, but these values also reflect differences in the difficulty of the items.

Psychometric properties of an item v_i can be described by the item characteristic curve, i.e. the chance $P_i(1|\theta)$ to master v_i given the ability level θ (see figure 1). It seems not unreasonable

Insert fugure 1 about here (see page 32)

(Lord, 1970) to assume that the item characteristic curves have the normal ogive form, i.e. they have the mathematical properties of a normal cumulative distribution (Assumption 2). See for a discussion of the normal ogive model Lord & Novick (1968). The normal ogive item characteristic curve of v_i has a point of inflection at $\theta = b_i$ (see figure 2). At this point of the θ-scale the probability of a correct answer is .50; b_i is called the difficulty of the item v_i

Insert figure 2 about here (see page 32)

Lord and Novick (1968, p. 375) show that from assumptions 1 and 2 follows that $b_i = {}^{z}i/\rho_i$, with $\rho_i \equiv$ biseral correlation between the dichotomized itemscore and θ, while z_i is defined by

$$\pi_i \equiv \int_{z_i}^{+\infty} \phi(\theta)\, d\theta$$

in which the stochastic variable π_i is estimated by p_i ($i = 1, 2, \ldots, m$); ρ_i is called the discriminating power of item v_i. For all the items within the analysis, we assume that the discriminating power ρ_i is the same, i.e. $\rho_i = \rho$ (Assumption 3). By this, ρ has become a scale factor which can be left out of consideration. Thus, we may write: $b_i = z_i$, in which z_i is

estimated via the above formula using p_i.

In $b_i=z_i$ we have for each item v_i a measure of difficulty on the same quantitative scale θ. Earlier in this section, we saw that the values p_i of the respective items v_i ($i=1, 2, \ldots, m$) not only are an estimation of a measure of attainment of the objective represented by each v_i, but also reflect the differences in difficulty of the respective items. Now, we can not only interpret z_i ($i=1, 2, \ldots, m$), estimated via p_i, as a measure of difficulty for item v_i, but we can also take the line that the z_i pertains to the attainment of the objective operationalized in the respective items v_i. This result will be used in the next section.

Assumption 3 can be tested by correcting, for each item v_i, the item-test correlation for attenuation and spuriousness and checking if these are the same for all items. However, this is not possible in the situation of a retrospective analysis because the students did not take all the items, since different examinations were taken by different students.

3.4. Measure for the attainment of objectives

In this section, measures for the attainment of objectives by the students will be discussed. We will discuss the cases of cells with only one, respectively at least two items. In each case the criterion for attainment of the objective has to be chosen; this is a matter of rational judgement.

a. Cells with 1 item.

At the end of § 3.2. a measure p of attainment for an objective operationalized in an item v was introduced. In this case, it is easily to define (without needing the assumptions of § 3.3.): the students have attained the course objective represented by item v if $p \geq \beta$ with $0 \leq \beta \leq 1$. For example we may judge that a given objective is attained if at least 75% of the students have reached the chosen level α of mastery of an item (thus $\beta = .75$).

b. Cells with more than 1 item.

Before discussing the two approaches, we shall define several notations with respect to a particular cell of the grid:

- items : $v_1, v_2, \ldots, v_m$
- number of students (generally different for each item) : $N_1, N_2, \ldots, N_m$
- number of students that master an item : $n_1, n_2, \ldots, n_m$
- measure of attainment of an item; $p_i = {}^{n}i/N_i$: $p_1, p_2, \ldots, p_m$

b1. Referring to the definition of p_i of each item v_i and to the case of a cell with 1 item, we can consider all the items which apply $p_i \geq \beta$. Let us denote by x the number of items from $v_1, v_2, \ldots, v_m$ for which the students have attained the chosen level of mastery α for each item. We define as follows

The group of students has attained the objective if $\frac{x}{m} \geq \gamma$, which $0 \leq \gamma \leq 1$.

The choice of γ is again by judgement. By choosing this definition of attainment, of the assumptions in § 3.3. we only need the first part of assumption 1, viz. that the combined influence of the sources of variation (a) students and (b) instruction is constant over the considered period of time.

Note that in this definition only the number of problems by which an objective is operationalized is considered. It is possible that the value of m could be very small, by which the application of this measure is not very meaningful. Example: If $m = 1$, then $\frac{x}{m}$ can be only 0 or 1. If $m = 2$, then $\frac{x}{m}$ can be only 0, ½ or 1.

By taking into account only the number of items, all items have equal weight in the procedure. Use is not made of the fact that different numbers of students are performing on the respective items. Thus, the score of a single student performing on an item taken by a small number of students has a relatively greater weight than in the case of an item taken by a greater number of students. This will not be a problem if (a) all the items are of almost the same difficulty (i.e. the values $p_1, p_2, \ldots, p_m$ are almost equal) or (b) the

numbers of students taking the respective items are almost equal.

In educational practice those conditions are not always fulfilled. An example will help to illustrate the point. Let an objective be operationalized in three examinations in three different items and let the number of students in these examinations be 300, 400 and 40 respectively. This is a realistic situation in our university. Moreover, let us assume that only one item is attained ($p \geq \beta$) by the group of students. If the attained item is the one made by the group of 400 students and the two not-attained items are those in the examinations with the smallest number of students than, considering the whole population of students it may be reasonable to regard the item presented to the large group as a more representative operationalization of the objective than the two other items.

From this point of view, it is desirable to look for a procedure in which not only the difficulty of the items, but also the number of students taking the items will considered.

b2. To arrive at a procedure for which the two conditions are fulfilled the assumptions of § 3.3. are needed. There, we saw that for each item v_i ($i = 1, 2, \ldots, m$) in a cell we have a z_i estimated by p_i via the transformation

$$p_i = \int_{z_i}^{+\infty} \phi(\theta) d\theta,$$

where $\phi(\theta)$ is the standard normal distribution. The z_i ($i = 1, 2, \ldots, m$) can be found in a table of the standard normal distribution. It was concluded that z_i is not only a measure of difficulty of item v_i, but also refers to the attainment of the objective operationalized by v_i.

By computing for each cell

$$z^* = \sum_{i=1}^{m} N_i z_i \Big/ \sum_{i=1}^{m} N_i$$

the different numbers of students per item are taken into account. Moreover, the weighted mean z^* of difficulties can be considered as representing the average difficulty of the items about the objective students are expected to attain. By means of the inverse transformation at each z^*, a value p^* can be found. This p^* can be interpreted as a sort of 'mean' value of the p_i's for the items in a cell. Note, that p^*, as the transformed of the weighted mean of z_i's, is not the same as the weighted mean of the p_i's.

For each objective, corresponding to a cell of the grid, p^* can be interpreted as a measure of the attainment of the objective. Therefore, we define:

<u>The group of students has attained the objective if $p^* \geq \beta$, with $0 \leq \beta \leq 1$.</u>

Consistent with other parameters, the choice of β is a matter of judgement.

Note, that the case of one item per cell is a special case (viz $m = 1$) of this.

4. Application

4.1. Available data.

We have analyzed the mathematics courses for first year engineering students at our university in the first three quarters of 1970/71, 1971/72 and 1972/73. Of each course the examinations at the end of the quarter, immediately after the teaching period, and at the end of the course year are analyzed.

In § 2 we pointed out that in this analysis the central question is: what objectives were in fact attained by the group of students which passed the examinations. Thus the method to determine the attainment of objectives was applied to this specific group.

The data for this retrospective analysis were collected in 1976. It appeared, that not every teacher of mathematics classes had saved the list with scores of the students on the respective problems. So, the analysis could only be applied on a part of the students that took the examinations. Table 1 contains the number of available data per examination.

Each examination was labeled with a number. The first digit indicates the quarter of the course year in which the mathematics course was given. The second digit indicates the several examinations of the particular course.

Insert table 1 about here (see page 33)

From table 1 we see that the available data on some examinations were very small. Therefore, the analysis of what objectives were pursued is based upon all the examinations, but the analysis of what objectives were attained is only applied on those examinations with at least 15 passed students.

In § 4.5., the measure for attainment of objectives discussed in § 3.4. under b2 is applied. With respect to the assumptions discussed in § 3.3., the following can be said. Assumptions 2 and 3 are underlying the analysis, their fulfilment was presupposed. The first part of assumption 1 is plausible for the considered mathematics courses in the period 1970-1973, because in this period there were no changes in the entry requirements, the preceeding instruction (secondary school), the group of executive teachers, the instructional procedures (lectures and practice sessions) the basic textbook, etc. Finally it was assumed for the group of passed students that the variable θ, expressing the ability to solve mathematics problems has a standard normal distribution.

From these considerations we conclude that the results of

the analysis will have no general validity. They must be considered mainly as illustrations of the possibilities of the method.

4.2. Grid of objectives.

The problems in the examinations where distributed over the following content categories:

1. Mathematical induction;
2. Limits (of sequences; functions of one variable: continuity, differentiability, computing of limits. theorem of l'Hopital);
3. Differential calculus (finding the derivative, linear approximation, mean value theorem, extrema, points of inflection);
4. Integral calculus (primitive function, Riemann sum, definite integral, length of arc, area and volume of surfaces of revolution);
5. Functions of two variables (continuity, differentiability, partial derivatives, linear approximation);
6. Differential equations (different types of first order equations);
7. Series (tests for convergence and divergence, power series, series expansions of functions, Taylor's formula, computation with series);
8. Improper integrals (tests for convergence and divergence)

These categories reflect the main topics in the three mathematics courses. The categories 1, 2 and 3 refer to the first quarter course: mathematics I , 4, 5 and 6 to mathematics II and 7 and 8 to mathematics III. The number of categories is such that each

item could be categorized unambigiously according to content, while yet each category contains a meaningful number of items.

Indicated in § 3.1. is the classification in levels of behaviour using the main categories of Poyla (1957), who distinguishes between "problems to find" (F) and "problems to prove" (P). Additionally a rest category, "reproduction" (R), was used.

F. Problems to find

The items in this category are characterized by "compute" "determine", "find" and/or "approximate". Three levels have been distinguished:

F1: elementary problems, solved with only one standard method (i.e. a for the students well known method);

F2: problems solved by using several standard methods after each other or by using a standard method after a non-trivial re-formulation of the problem;

F3: problems solved with non-standard methods (i.e. new problems).

P. Problems to solve

These items are characterized by words as: "prove" or "examine if ...". We distinguish four levels:

P1: problems indicating which definition or theorem must be applied and/or which standard type of proof can be used;

P2: problems wherein the theorem to be used is not given and/ or a non-trivial re-formulation of the problem is necessary as a first step;

P3: problems in which several definitions, theorems and/or standard types of proof must be used (i.e. new problems);

P4: proofs of generalizations or specializations of well-known theorems, of which it can be expected that the students have not learned them by heart.

R. Reproduction

This category is not further subdivided.

4.3. Categorization of problems

To indicate how we have interpreted both type of categories of the preceeding section, we offer some examples in this section. It appeared that it was not always easy to put the items in a behavioral category. Sometimes it was difficult to distinguish between the categories F1 and P1; e.g. the differences between problems to prove and problems to find on 7 series fade away because the use of a test for convergence or divergence is in fact the computation of a standard limit. Or: finding an interval of convergence of a power series is in fact fully analogous to examining the convergence of a series with fixed terms.

4.3.1. Examples of problems to find.

F1: integral calculus; p = .85

Find:

$$\int \frac{1+x^2}{x^2(1-x)}\,dx.$$

F2: differential equations; p = .63

Solve:

$$y\frac{dy}{dx} = y^2 + xy^2 + x + 1.$$

F1: series; p = .86

Find the interval of convergence of

$$\sum_{n=2}^{\infty} \frac{nx^n}{\ln x}.$$

F2: limits; p = .76

Given the function f with

$$f(x) = \frac{\ln x}{1+\ln^2 x},\ x>0.$$

Define f in the point 0 so that f has right hand continuity in that point.

F2: differential calculus; p = .64

Given the function f with

$$f(x) = \frac{2}{x-1} + \frac{1}{2}\ln (x+1)^2.$$

Find the extrema, points of inflection and asymptotes; find the intervals on which f is increasing and decreasing and also the intervals on which the graph of f is concave upward and concave downward (using the second derivative method). Sketch the graph of f.

4.3.2. Examples of problems to prove.

P1: functions of two variables; p = 0.33

Prove:

$$\lim_{\substack{x\to 0\\ y\to 0}} \frac{x^2y^2}{x^2+y^2} = 0$$

P1: series; p = .86.

Examine the following series for (absolute) convergence or divergence:

$$\sum_{n=1}^{\infty} (-1)^{n-1} \frac{n^2+3n+5}{2n^2+4n-1}$$

P2: limits; p = .25.

Given the function f with

$$f(x) = x^{-x \ln x}, \quad x > 0$$

$$= 0 \qquad , \quad x \quad 0$$

Prove that

$$\lim_{x\to 0^+} f(x) = 0;$$

find the right hand derivative in 0

P2. series; p = .25.

The n-th term of a series is given by

$$t_n = \int_n^{n+1} \frac{dx}{x^3+1}.$$

Prove the convergence of

$$\sum_{n=0}^{\infty} t_n.$$

P4: differential calculus; p = .35.

A function f is three times continuous differentiable with f(0) = 1, f(1) = 2 and f'(0) = 1. Prove, using the mean value theorem, that there is a point a with $0 < a < 1$ such that $f''(a) = 0$.

4.4. Pursued course objectives

The result of categorizing the examination items within the grid is summarized in table 2

Insert table 2 about here (see page 34)

By analyzing information in this table, some important conclusions pertaining the pursued objectives can be derived:

a. Apparently, in examinations the mathematics teachers laid stress on an operationalization of the objectives mainly on levels F1, F2 and P1 and, to less extent, P2. The remaining categories ("new problems") did not or seldom appear.

b. The grading of mathematics examinations uses compensatoric model, i.e. a student gets a pass mark if his examination score is at least 55% of the maximum score. From table 2 we concluded that a student could receive a pass mark if he disregarded preparation for problems on the level of P2, P3 and P4. This means that while higher level problems were indeed discussed during the lectures and discussion groups, the department of mathematics accepted that students needed only take cognizance of the higher level objectives, but that they did not need to master them.

c. Less attention was paid to "problems to prove" then to "problems to find". Exceptions were the content categories 7 "series" and 8 "improper integrals". But in section 4.3. we have pointed out the difficulty with the categorization of the problems on series;

since the relationship between the problems on improper integrals and the indicated problems on series, the same held to the category improper integrals.

d. The row totals reflect the relative importance of the respective content categories. Two comments have to be made: (a) category 1: mathematical induction was indeed a small topic, only treated on the level P1 and (b) the numbers in categories 7 and 8 were based upon four examinations, the remaining upon three examinations.

4.5. Attainment of course objectives

In table 3 the computations according to the method b2 in § 3.4. are summarized. The results are based on a portion of the examinations (see table 1). The raw data are presented in the appendix.

Before discussing conclusions from an analysis of table 3, some parameters have to be discussed.

The choice of the parameter α for mastery of a problem (see 3.2.) by a student was $\alpha=.66$. The rationale behind this choice is the following. Most of the essay problems in the examinations had a maximum score between 2 and 5 points. A student who made a minor error, e.g. in the computation or in the wording of the solution, would not get the maximum score. However, we could assume he/she understood the solution of the problem.

In this case the score on a problem was often 1½ out of 2 points 2 out of 3 points, 3 out of 4 points, etc. i.e. a score of at least .66 of the maximum score.

In § 3.4. we discuss two measures to express whether the group of students has attained its goals or not. For reasons explained in § 3.4. we will choose p^* , defined in that section, as measure for the attainment of the goals. This measure expresses how well the group of students, in this paper the group of passed students, has attained the course objectives.

When have the passed students attained the objectives? From table 2 we have concluded, that in the mathematics courses primarily objectives on the levels F1, F2 and P1 were pursued, and, to a less extent, level P2. The other levels were so scarcely represented in the examinations that earlier on is concluded that those levels were not pursued. Restricting ourselves, therefore, to the levels F1, F2, P1 and P2, we chose as criterion for attainment of an objective $\beta = .75$. Thus, if in a cell $p^* \geq .75$ was recorded we said that the objective was attained by the group of passed students.

Now, from table 3 some important conclusions were drawn pertaining to the attainment of the objectives by the group of passed students.

Insert table 3 about here on separate page (see page 35

---

a. All the objectives, but one (viz. (3, F2)), which were attained by the group of passed students ($p^* \geq .75$) appeared to be of the type F1: elementary standard problems to find. Looking at the columns, only the group of objectives F1 was attained by the students; although one cell was clearly not attained.

b. None of the objectives on the level of P1 and P2 was attained, although some cells were close to the criterion.

c. Because the row totals in table 3 were computed over different behavioral levels, it was not useful to apply the criterion of attainment of an objective on the row totals. From the last columm of table 3 can be seen that there were poor results on 1. mathematical induction, 2. limits and 5. functions of two variables.

d. Because the F1-type problems form usually less than half the number of items of an examination an interesting conclusion could be drawn with respect to the considered group of students. Although the passed students performed well on the attained problems of type F1, their score on these problems was usually not sufficient for getting a pass mark on the examinarion. To get the pass mark on the examination of the course every student has to master several of the objectives of which we concluded that they were not attained by the whole group of passed students. This means that teachers of following courses must take into account the fact that the group of students had attained only a relatively small part of the course objectives for courses which are prerequisites for his/here courses. This means that apparently it is the practice of mathematics education within the Department of Mathematics to accept passed students who may not have attained most of the course objectives but only taken cognizance of them.

5. Discussion

The procedure for a retrospective analysis of course objectives is not a test in a statistical sence. It is meant as an attempt to bring some order in an unsurveyable amount of examination results, so that discussions on these results will be possible. Then, discussions on the objectives of courses can be conducted using arguments which are based more upon objective data then personal feelings and impressions of teachers.

In the preceeding, we have discussed the questions: (a) which objectives were pursued in the mathematics courses and (b) which of them were attained by the students with a pass mark. The answers to the first question are important for improving the considered mathematics courses. The answers to the second question are important in connection with the construction of following courses of which the concerned course is one of the prerequisites. With this procedure we can indicate a general knowledge base upon which teachers of following courses may build upon.

We have considered the passed students of every examination. No distinction was made between examinations immediately after the teaching of the course and examinations later on in the year (a "second opportunity"), because the students who failed the examinations immediately after the course were provided an extra opportunity. Therefore, in our choice of the subgroup of students, the "passed students", no one was counted twice.

Some concluding remarks and points for further research have to be mentioned.

1. Under the assumptions presented in § 3.3. per cell, the measure of attainment p^* of an objective is computed in § 3.4. via the transformation $z_i^* \rightarrow p_i^*$. The transformations are carried out via the cumulative standard normal distribution. From a table of this distribution can be seen that p varies alsmost proportionally with z outside the trials of the distribution. Thus, to get a quick impression of the results the weighted means of p_i-values can be computed, as long as the p_i's are not too small or too large.
2. The number of items per cell will influence the accuracy of the estimation by p^* of the true value of the measure of attainment of an objective. How to bring this element into the method discussed in § 3.4. has to be studied.
3. The method presented in this paper can also be applied to examine the quality of instruction. Then, the measures of attainment of an objective have to be interpreted as measures for the quality of instruction. The examinations immediately after teaching of the course have to be analyzed for all the students who took them.

Acknowledgement

We owe many thanks to Dr. Wim J. van der Linden and Dr. Robert C. Harris for their willingness to contribute to this paper by intensive discussion of the theoretical part and critical comment on the manuscript.

References

Bloom, B.S. (ed.), Taxonomy of educational objectives: the classification of educational goals; Handbook I: Cognitive domain. New York: McKay, 1956.

Lord, F.M., Item characteristic curves estimated without knowledge of their mathematical form - A confrontation of Birnbaum's logistic model. Psychometrika, 1970, 35, 43-50.

Lord, F.M., and Novick, M.R., Statistical theories of mental test scores. Reading (Mass): Addison-Wesley, 1968.

Polya, G., How to solve it. Garden City (N.Y.): Doubleday, 1957.

Wilson, J.W., Evaluation of learning in secondary school mathematics. In: B.S. Bloom et al., Handbook on formative and summative evaluation of student learning. New York, McGraw-Hill, 1971.

Appendix

This appendix contains the data upon which the retrospective analysis of the attainment of objectives is conducted. The data are grouped per content category. The colums are consisting of:

a. the behavioral levels F1, P1, etc.

b. per item v_i the number N_i of students that passed the examination to which the item belongs.

c. per item v_i the number n_i of students that masters the resp. items

From these data table 3 can be derived: the number of items can be counted, per item $p_i = {}^{n}i/N_i$ can be computed, etc.

1. Mathematical Induction		
a	b	c
P1	40	9
	17	11
	13	8
	28	16
2. Limits		
a	b	c
F1	40	14
	17	8
	28	9
	17	7
	17	13
	28	22
F2	17	11
	28	15
	17	13
	14	19
	17	9
	17	8

Limits (cont.)		
a	b	c
P1	28	15
	17	6
	28	17
	28	5
P2	17	3
	40	10
P3	17	0
P4	17	4
3. Differential Calculus		
a	b	c
F1	40	38
	17	14
	17	11
	17	14
	17	11
	17	16

Differential Calculus (cont.)		
a	b	c
F1	17	13
(cont.)	28	27
	17	15
F2	17	12
	28	26
	40	38
	40	21
	17	14
	28	18
P1	28	12
P4	40	14
	17	8
4. Integral Calculus		
a	b	c
F1	18	14
	14	6

Integral Calculus (cont.)		
a	b	c
F1	53	44
(cont.)	18	15
	18	13
	18	13
	18	8
	18	16
	53	45
	53	49
	18	15
F2	53	11
	53	31
	53	24
	18	14
	18	11
	78	5
	53	29
	18	18

Integral Calculus (cont.)		
a	b	c
P1	18	15
	18	5
	18	10

5. Functions of two variables		
a	b	c
F1	18	15
P1	18	5
P2	18	6
	18	3

6. Differential Equations		
a	b	c
F1	18	15
	53	36

Differential Equations (cont.)		
a	b	c
F2	53	39
	18	12
	18	16
	18	10
P1	18	5

7. Series		
a	b	c
F1	50	44
	38	27
	87	75
	87	55
	87	72
	72	9
	50	44
	50	33
	38	31
	36	33
	38	36

Series (cont.)		
a	b	c
F1	50	42
(cont.)	87	75
	36	21
F2	50	10
	42	10
P1	50	31
	38	33
	87	54
	38	30
	87	75
	50	41
	50	35
	38	29
	38	12
	87	62
P2	87	62
	87	52
	50	35

Series (cont.)		
a	b	c
P4	50	10
	38	16
	38	13

8. Improper Integrals		
a	b	c
F1	38	37
P1	87	61
	50	47
	50	29
	50	21
	38	18
	38	28
P2	50	21
	38	6

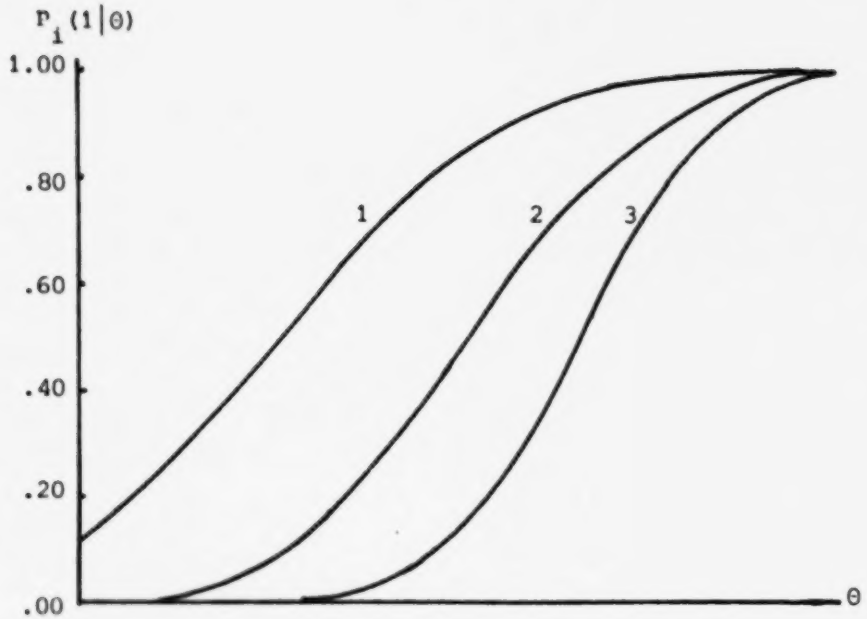

Fig. 1: Item characteristic curves

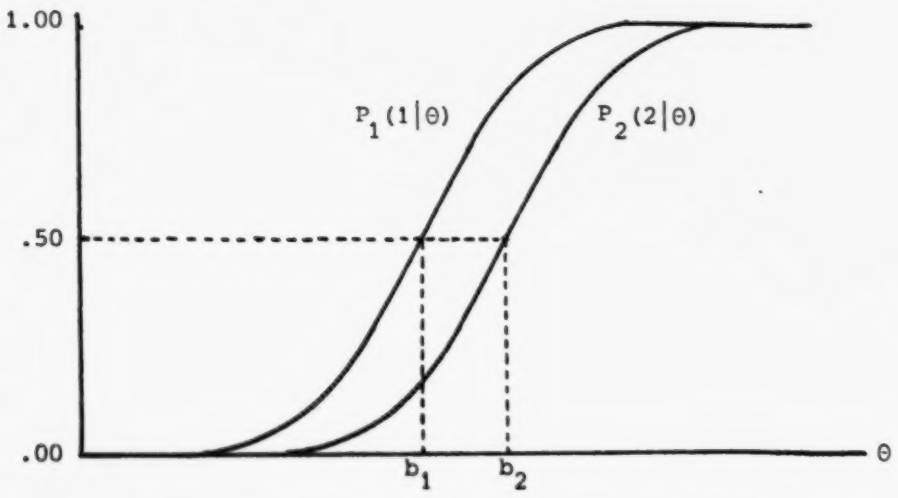

Fig. 2: Normal ogive item characteristic curves.

Tabel 1: Number of students per examination

exam. number	11*	12*	13*	14	15*	16	21*	22	23*	24	25*	26	31*	32	33	34	35*	36	37*
number of students	66	21	40	16	46	4	85	12	37	10	46	21	107	4	38	23	104	31	55
passed students	40	17	17	12	28	3	53	5	18	8	18	7	87	3	10	13	50	14	38

*) These examinations are used in the analysis of attained objectives.

Table 2: Number of items per objective.

	Problems to find				Problems to prove					Reprod.	Total
	F1	F2	F3	Total F	P1	P2	P3	P4	Total P		
1. mathematical induction	-	-	-	-	4	-	-	-	4	-	4
2. limits	9	11	-	20	7	4	1	1	13	3	36
3. differential calculus	16	10	-	26	1	1	-	4	6	-	32
4. integral calculus	18	15	-	33	3	-	-	-	3	-	36
5. functions of two variables	6	-	-	6	2	6	-	-	8	-	14
6. differential equations	7	7	-	14	1	-	-	-	1	-	15
7. series	28	7	-	35	20	10	1	4	35	-	70
8. improper integrals	2	1	-	3	10	2	-	-	12	-	15
Total	86	51	-	137	48	23	2	9	82	3	222

Table 3: Results of the retrospective analysis

	F1		F2		P1		P2		P3		P4		Total [3]	
	m [1]	p* [2]	m	p*	m	p*	m	p*	m	p*	m	p*	m	p*
1. mathematical induction	-	-	-	-	4	.44	-	-	-	-	-	-	4	.44
2. limits	6	.50	6	.55	4	.41	2	.23	1	.00	1	.24	18	.46
3. differential calculus	9	.88	6	.80	1	.43	-	-	-	.00	2	.39	16	.82
4. integral calculus	11	.82	8	.56	3	.57	-	-	-	-	-	-	22	.69
5. functions of two variables	1	.83	-	-	1	.28	2	.24	-	-	-	-	4	.40
6. differential equations	2	.72	4	.73	1	.59	-	-	-	-	-	-	7	.69
7. series	14	.77	2	.22	10	.73	3	.67	-	-	-	-	29	.71
8. improper integrals	1	.97	-	-	6	.68	2	.10	-	-	-	-	9	.67
Total	45	.78	26	.60	30	.65	9	.48						

1) m is the number of problems in which the objective is operationalized.

2) p^* the measure for attainment of the objectives by the group of passed students (see § 3.4.)

3) only based upon the F1, F2, P1 and P2 columns.

REEDS VERSCHENEN TWENTS ONDERWIJSKUNDIGE RAPPORTEN

1. Linden, W.J. van, & Mellenbergh G.J. Coefficients for tests from a decision theoretic point of view, juni 1977.

2. Linden, W.J. van, & Mellenberg, G.J. The linear Utility Model for Optimal Selection, oktober 1977.

3. Linden, W.J. van, Forgetting, Guessing, and Mastery: The Macready and Dayton models revisited and compared with a latent trait approach, december 1977.

4. Linden, W.J. van, & Mellenberg G.J. The Internal and External Optimality of Decisions based on Tests, maart 1978.

5. Linden, W.J. van, Binomial Test Models and Item Difficulty, mei 1978.

6. Diana, I.P.F. de, Evaluatie van beroepenoriëntatie in het participatie-onderwijs: een onderzoek naar het verloop en de effekten van een programma beroepenoriëntatie, juli 1978.

REEDS VERSCHENEN TWENTS ONDERWIJSKUNDIGE MEMORANDA

1. Plomp, Tj. Bijzondere differentiatiegebieden: een analogon van de technieken van Johnson-Neyman en Potthoff, mei 1978.

2. Krammer, H.P.M. Een mathematisch model van het leren van 'vaag gedefineerde' begrippen, mei 1978.

3. Warries, E. Studiekeuze en selectie: differentiële onderwijseffekten voor vrouwen en mannen, mei 1978.

4. Linden, W.J. van, Het Klassieke Testmodel, Latente Trekmodellen en Evaluatieonderzoek, juli 1978.

5. Krammer, H.P.M. Onderwijskundige principes voor de ontwikkeling van een Lerarenopleiding, juni 1978.

6. Tomic, W. Puberteit, augustus 1978.

7. Tomic, W. Ontwikkeling van de intelligentie, oktober 1978.

8. Warries, E. Het produceren van leermateriaal, oktober 1978

9. Warries, E. Opmerkingen over selectieve toelating, oktober 1978

10. Diana, I.P.F. de, & Stuart, M.J.W. Formatieve evaluatie van de ontwikkeling van een Leerplan Algemene Techniek in een integraal middenschool experiment, december 1978.